

526,000

Rec'd PCT/PTC 24 FEB 2005

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international(43) Date de la publication internationale
11 mars 2004 (11.03.2004)

PCT

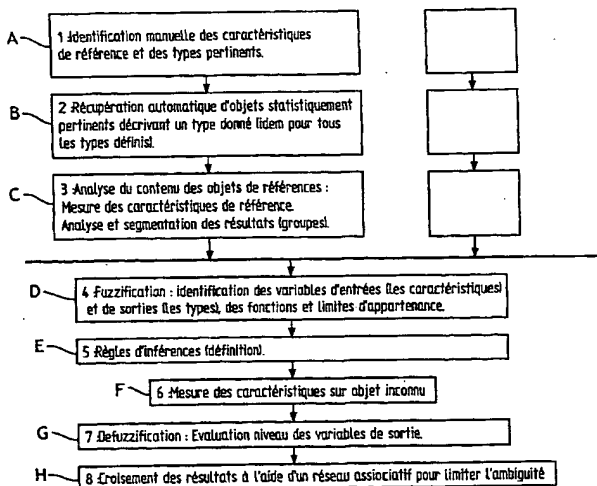
(10) Numéro de publication internationale
WO 2004/021265 A2

- (51) Classification internationale des brevets⁷ : G06K 9/62
- (21) Numéro de la demande internationale : PCT/FR2003/002588
- (22) Date de dépôt international : 27 août 2003 (27.08.2003)
- (25) Langue de dépôt : français
- (26) Langue de publication : français
- (30) Données relatives à la priorité :
02/10776 30 août 2002 (30.08.2002) FR
- (71) Déposant (pour tous les États désignés sauf US) :
FRANCE TELECOM [FR/FR]; 6, place d'Alleray,
F-75015 Paris (FR).
- (72) Inventeur; et
- (75) Inventeur/Déposant (pour US seulement) : LANCIERI,
Luigi [FR/FR]; 16, rue Jacques Brel, F-14210 Evreux (FR).
- (74) Mandataires : MARTIN Jean-Jacques etc.; Cabinet
Regimbeau, 20, rue de Chazelles, F-75847 Paris Cedex 17
(FR).
- (81) États désignés (national) : AE, AG, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ,
DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD,
SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, YU, ZA, ZM, ZW.

[Suite sur la page suivante]

(54) Title: FUZZY ASSOCIATIVE SYSTEM FOR MULTIMEDIA OBJECT DESCRIPTION

(54) Titre : SystEme associatif flou de description d'objets multimEdia



A1...MANUALLY IDENTIFYING REFERENCE CHARACTERISTICS AND RELEVANT TYPES
 B2...AUTOMATICALLY RETRIEVING STATISTICALLY RELEVANT OBJECTS DESCRIBING A GIVEN TYPE (SAME FOR ALL DEFINED TYPES)
 C3...ANALYZING CONTENT OF REFERENCE OBJECTS: MEASURING REFERENCE CHARACTERISTICS. ANALYZING AND SEGMENTING RESULTS (GROUPS)
 D4...FUZZIFYING: IDENTIFYING INPUT VARIABLES (CHARACTERISTICS) AND OUTPUT VARIABLES (TYPES), FUNCTIONS AND MEMBERSHIP LIMITATIONS
 E5...INTERFERENCE RULES (DEFINITION)
 F6...MEASURING CHARACTERISTICS ON UNKNOWN OBJECTS
 G7...DEFUZZIFYING: EVALUATING LEVEL OF OUTPUT VARIABLES
 H8...CROSS-CHECKING RESULTS BY MEANS OF AN ASSOCIATIVE NETWORK TO LIMIT AMBIGUITY

(57) Abstract: The invention concerns a method for automatic description of an unknown multimedia object, which consists in associating the unknown object with several types of reference multimedia objects each time based on a membership probability particular to each type concerned (G), the method including a step which consists in measuring at least one physical characteristic on the unknown object (F) and in comparing it with measurements of characteristics representing best the reference types, the method further comprising the step (H) which consists in using for each type at least one probabilistic relationship providing a membership probability to the type concerned based on the result of the comparison of the characteristics of the unknown object and of the type, the method further comprising the step which consists in using the probabilities of membership to the various types thus obtained in combination with a series of affinity relationships between types, so as to select memberships which both constitute the majority in terms of probability and which designate one another by their affinity relationship, and so as to exclude memberships with less affinity with the selected types.

[Suite sur la page suivante]

WO 2004/021265 A2



(84) États désignés (*régional*) : brevet ARIPO (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Déclaration en vertu de la règle 4.17 :

— *relative à la qualité d'inventeur (règle 4.17.iv)) pour US seulement*

Publiée :

— *sans rapport de recherche internationale, sera republiée dès réception de ce rapport*

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

(57) **Abrégé :** L'invention concerne un procédé de description automatique d'un objet multimédia inconnu, dans lequel on associe l'objet inconnu avec plusieurs types d'objets multimédia de référence selon à chaque fois une probabilité d'appartenance respective à chaque type considéré (G), le procédé incluant une étape consistant à mesurer au moins une caractéristique physique sur l'objet inconnu (F) et à la comparer avec des mesures de caractéristiques représentant au mieux les types de référence, le procédé comprenant en outre l'étape (H) consistant à utiliser pour chaque type au moins une relation probabiliste donnant une probabilité d'appartenance au type considéré en fonction du résultat de la comparaison des caractéristiques de l'objet inconnu et du type, le procédé comprenant en outre l'étape consistant à utiliser les probabilités d'appartenance aux différents types ainsi obtenues en combinaison avec une série de relations d'affinités entre types, de manière à élire des appartenances qui sont à la fois majoritaires en probabilité et qui se co-désignent par leur relation d'affinité, et de manière à exclure des appartenances ayant une moindre affinité avec les types élus.

« Système associatif flou de description d'objets multimédia »

L'invention concerne la caractérisation d'objets multimédia, et en
5 particulier les techniques pour réduire les ambiguïtés de la reconnaissance
d'objets multimédia.

L'invention concerne ainsi, notamment, la description du contenu des
fichiers multimédia associés aux objets.

L'objectif de l'invention est notamment de proposer un procédé
10 permettant de décrire et de comparer des objets dans un contexte
d'incertitude.

Pratiquement, l'invention vise par exemple à décrire et comparer des
fichiers image, musicaux, vidéo ou d'autres formes.

Elle permet de répondre aux questions du type : je souhaite trouver
15 des morceaux de musiques correspondant à « mes goûts ». Une telle
opération pose deux problèmes. Le premier est de modéliser des notions
comme « mes goûts » qui sont très subjectives. De nombreux travaux ont
déjà traité ce problème dans le cas où l'objet bénéficie d'un contexte
textuel. Nous nous intéressons notamment ici aux objets isolés sans
20 contexte textuel. C'est la deuxième difficulté que nous proposons de
résoudre, à savoir extraire d'un objet isolé des caractéristiques qui
permettront de le décrire et de le soumettre à la comparaison, l'objectif
étant d'en obtenir une caractérisation sémantique (e.g. sous forme de mots
clés) en termes de probabilité.

25 Ces objectifs sont atteints dans le cadre de l'invention, grâce à un
procédé de description automatique d'un objet multimédia inconnu, dans
lequel on associe l'objet inconnu avec plusieurs types d'objets multimédia
de référence selon à chaque fois une probabilité d'appartenance respective
à chaque type considéré, le procédé incluant une étape consistant à
30 mesurer au moins une caractéristique physique sur l'objet inconnu et à la
comparer avec des mesures de caractéristiques représentant au mieux les
types de référence, le procédé comprenant en outre l'étape consistant à
utiliser pour chaque type au moins une relation probabiliste donnant une

probabilité d'appartenance au type considéré en fonction du résultat de la comparaison des caractéristiques de l'objet inconnu et du type, le procédé comprenant en outre l'étape consistant à utiliser les probabilités d'appartenance aux différents types ainsi obtenues en combinaison avec
5 une série de relations d'affinités entre types, de manière à élire des appartenances qui sont à la fois majoritaires en probabilité et qui se co-désignent par leur relation d'affinité, et de manière à exclure des appartenances ayant une moindre affinité avec les types élus.

On propose également selon l'invention un dispositif de description
10 automatique d'un objet multimédia inconnu, comprenant des moyens pour associer l'objet inconnu avec plusieurs types d'objets multimédia de référence selon à chaque fois une probabilité d'appartenance respective au type considéré, le dispositif incluant des moyens pour mesurer au moins
15 une caractéristique physique sur l'objet inconnu et la comparer avec des mesures représentant au mieux les types de référence, le dispositif comprenant en outre des moyens utilisant pour chaque type au moins une relation probabiliste donnant une probabilité d'appartenance au type
20 considéré en fonction du résultat de la comparaison des caractéristiques de l'objet inconnu et du type, le dispositif comprenant en outre des moyens pour utiliser les probabilités d'appartenance aux différents types ainsi
obtenues en combinaison avec une série de relations d'affinités entre types, de manière à élire des appartenances qui sont à la fois majoritaires
en probabilité et qui se co-désignent par leur relation d'affinité, et de
25 manière à exclure des appartenances ayant une moindre affinité avec les types élus.

D'autres caractéristiques, buts et avantages de l'invention apparaîtront à la lecture de la description détaillée qui va suivre, faite en référence aux figures annexées sur lesquelles :

- les figures 1a et 1b représentent schématiquement des modes
30 d'association, le premier mode dans un plan dit « de types », le second mode en tant que mode à apprentissage,

- la figure 2 est un synoptique représentant différentes étapes mises en œuvre dans une variante préférée de l'invention,

- la figure 3 est un tracé représentant un niveau de probabilité d'appartenance en fonction du niveau de pondération d'une caractéristique physique donnée,

- la figure 4 est un tracé représentant une approximation du tracé précédent,

- la figure 5 est un tracé représentant une probabilité d'appartenance en fonction d'une erreur mesurée sur des variables d'entrée,

- la figure 6 est un tracé du même type pour une erreur sur la variable de sortie,

- la figure 7 et la figure 8 sont chacune un tracé représentant une probabilité d'appartenance en fonction d'une erreur sur une caractéristique respective en entrée (étape 4 de la figure 2),

- la figure 9 est un tracé représentant une probabilité d'appartenance en fonction d'une moyenne d'erreurs mesurées sur deux caractéristiques.

La méthode employée est d'une grande adaptabilité. Trois aspects sont développés ci-après, qui sont les réseaux associatifs (liens sémantiques entre objets de références), la logique floue (gestion de l'incertitude) et une étape appelée ici « inférence corrélatrice » (limiter l'ambiguïté par croisement d'informations).

Contrairement à l'approche courante qui se place très en amont et cherche plutôt à optimiser la précision de la modélisation de l'image, plus généralement de l'objet, on se situe plus haut dans les couches. On prend le modèle de l'image (du son, etc.) brut même s'il est de mauvaise qualité et on cherche à réduire l'ambiguïté de l'interprétation grâce à des méthodes de gestion de l'incertitude (recoupement, logique floue).

L'opération de description est préférentiellement réalisée en comparant des objets inconnus à des objets de caractéristiques connues et par croisement et aboutit à un vecteur de caractéristique associé à une probabilité d'appartenance. Ce vecteur traduit la relation entre des propriétés physiques ou logiques et des propriétés conceptuelles (voir fig. 1a, 1b).

La méthode décrite ci-après vise la caractérisation dans un contexte d'incertitude et s'applique à des objets quelle que soit leur nature

(analogique, numérique, symbolique, etc.). Ceci étant dit, pour des raisons de clarté, nous restreignons les explications qui suivent aux fichiers de type multimédia (e.g. mp3, mpeg, wav, jpg, etc). Ces objets peuvent être par exemple décrits par certaines des propriétés physiques des objets analogiques sous jacents (fréquences, durée, formes typiques, etc.). Par ailleurs, ces mêmes objets peuvent être décrits dans le registre conceptuel par une typologie adéquate (musique pop, jazz, film d'action, photo de voiture) ou un type préféré sans qu'il soit référencé précisément (eg un mélange de jazz et de classique personnalisé). Un premier objectif est de mettre en œuvre une correspondance associative entre le domaine physique et le domaine conceptuel.

Les figures 1a et 1b présentent deux modes d'associations. Le premier (figure 1a) dans le plan des types est une association conceptuelle classique (par exemple entre des thèmes du lexique interconnectés par des relations pondérées en fonction de leurs proximités sémantique). Le second (figure 1b) est obtenu par apprentissage entre les types et les caractéristiques de références. Ce type d'architecture permet par recoupement d'identifier des relations conceptuelles entre caractéristiques physiques ou logiques.

Par exemple, les types peuvent être des mots, les caractéristiques peuvent être des formes dans le cas d'images.

Le choix des caractéristiques est à réaliser préférentiellement une fois pour toute pour un ensemble de types donnés mais requiert une expertise. Par exemple un musicien et un électronicien seront à même d'identifier les grandeurs représentatives de la typologie musicale (par exemple la fréquence moyenne des 10 premières secondes d'un morceau de musique permet d'identifier une musique d'ambiance d'un morceau de hard rock). Dans les exemples qui suivent nous n'utiliserons que des caractéristiques simplifiées pour décrire notre méthode.

Voici un premier exemple correspondant au cas de documents sonores, permettant de comprendre le principe.

Pour savoir si un fichier musical contient un morceau de flûte ou de cor (2 types conceptuels simples) il faut pouvoir accéder au signal sonore codé dans le fichier (e.g. wav, mp3).

Ceci est réalisé par des algorithmes de décodage classiques. Le
5 signal sera analysé avec des outils de traitement du signal adéquats qui pourront déterminer sa fréquence (caractéristique physique simplifiée pour l'exemple). Si on considère des fichiers de référence (plusieurs fichiers du type cor, plusieurs fichiers du type flûte), on pourra déterminer leurs fréquence moyenne dans le cas du cor et dans le cas de la flûte. On pourra
10 constater que la fréquence $A + \Delta f$ est celle de la flûte et la fréquence $B + \Delta f$ celle du cor. Ces fréquences ainsi que les types associés seront le modèle de référence obtenu par une étude statistique sur un échantillon représentatif. En calculant la fréquence f_i d'un certain nombre de fichiers contenant des sons d'instruments inconnus, on pourra déterminer leur
15 appartenance à la catégorie flûte ou cor en fonction de la proximité (si $f_i - A <$ seuil alors i proche de flûte). Il est également envisagé que l'ensemble des fichiers de référence de type « cor » puisse contenir des fichiers ne contenant pas du cor. La fréquence moyenne contient alors intrinsèquement une marge d'erreur et sa proximité avec la fréquence d'un
20 fichier à analyser est analysée de manière probable, également pour tenir compte du fait que le calcul de la moyenne est en fait le calcul d'une moyenne probable (espérance mathématique).

Cet exemple est très simplifié car un fichier sonore (c'est encore plus compliqué pour les fichiers vidéo) est le plus souvent une agrégation
25 complexe de fréquences différentes dans le temps et dans l'espace.

Pour résoudre ce problème de complexité, on identifie préférentiellement un certain nombre de caractéristiques fréquentielles spatiales et temporelles (fixes et évolutives) qui seront à même de décrire en théorie n'importe quel signal. Voici un autre exemple plus précis d'une
30 définition de telles caractéristiques. L'analyse spectrale du signal sur toute sa durée permet de récupérer le taux de fréquence hautes et basses. Nous ne considérons que 10 plages de fréquences (f_1 à f_{10}) réparties sur le spectre audible (20 Hz à 20 KHz) par exemple $df_1 = 20 \text{ Hz} - 500 \text{ Hz}$,

- df2=500Hz-2 KHz, df3=2kHz-5KHz etc. Dans la pratique le découpage fréquentiel tiendra compte de la fréquence fondamentale des principaux instruments. Cette analyse peut être réalisée à différentes périodes du signal. Par exemple, nous calculons la distribution fréquentielle sur des
- 5 espaces de dix secondes espacés toute les trente secondes sur un signal d'une durée de trois minutes. Les six ensembles de dix plages fréquentielles (que nous appellerons échantillons) seront considérés comme les caractéristiques physiques d'un signal du morceau de musique. Si comme dans l'exemple précédent on connaît un morceau de musique de
- 10 référence « qui nous plaît » (sans forcément bien savoir le décrire sur le plan conceptuel – c'est peut être du jazz ou du blues, etc-). On peut calculer l'échantillon de référence et par la suite on peut appliquer à tout morceau de musique le calcul de l'échantillon que l'on comparera à l'échantillon de référence). En fonction de la proximité physique, on pourra
- 15 déduire la proximité conceptuelle vis à vis de ce type « qui nous plaît ». Autrement dit, plus les caractéristiques spatiales, fréquentielles et temporelles du signal à classer et celles du signal de référence seront proches, plus on sera certain que le morceau à classer est un morceau de ce type « qui nous plaît ».
- 20 Dans le cas où on a défini plusieurs types de morceaux qui nous plaisent, on calcule la proximité des caractéristiques du morceau inconnu vis à vis des caractéristiques de chacun de ces types.
- On obtient alors une série de probabilités d'appartenance du morceau inconnu pour chacun de ces types. Les types de morceaux qui
- 25 nous plaisent ont eux-mêmes des compatibilités ou affinités. Ainsi, certains de ces types sont des types musicalement proches en termes de perception, tels que par exemple un type « rock » ou un type « hard-rock ».
- D'autres types sont incompatibles tels qu'un type « musique de chambre » et un type « musique électronique ».
- 30 On utilise ce réseau d'affinités de perception de l'oreille entre les différents types, ainsi que les probabilités d'appartenance aux différents types, pour déduire une appartenance à une catégorie plus large

regroupant une partie seulement de ces types qui nous plaisent. On choisit ici d'appeler cette étape une « inférence corrélative ».

Cette appartenance à une catégorie plus large finalise la description automatisée de l'objet multimédia au départ inconnu.

- 5 En d'autres termes, on utilise les probabilités d'appartenance aux différents types ainsi obtenues en combinaison avec une série de relations d'affinités entre types, affinités représentatives de la perception humaine, de manière à élire des appartenances qui sont à la fois majoritaires en probabilité et qui se co-désignent par leur relation d'affinité, et de manière à
- 10 exclure des appartenances ayant une moindre affinité avec les types élus.

Les élections et exclusions dont on parle se concrétisent dans le cas présent par le fait d'identifier une partie seulement (une « catégorie » ici) des types qui nous plaisent.

- De même, si l'on utilise plusieurs morceaux de références « qui nous
- 15 plaisent » pour calculer les caractéristiques moyennes, on réalise un apprentissage des caractéristiques de références. Ces caractéristiques de référence sont une moyenne pouvant intégrer des morceaux donnés par erreur et donc cette moyenne est une référence « la plus probable », et la proximité d'un morceau considéré vis à vis de ces caractéristiques, est
- 20 utilisée pour donner une « probabilité » que ce morceau nous plaise. Plus le morceau est proche de la référence statistique, plus la probabilité qu'il nous plaise est grande. La meilleure probabilité qu'il nous plaise correspond à une totale similitude entre caractéristiques du morceau considéré et moyenne statistique des caractéristiques de morceaux de référence.

- 25 Dans le cas de documents visuels (image et vidéo), la caractérisation des documents visuels est préférentiellement basée sur le même principe que celui des documents sonores. Il s'agit de mesurer la proximité avec d'autres objets selon des caractéristiques simples. Une image est analysée pour identifier par une recherche de contours les formes complexes isolées
- 30 ainsi que leurs caractéristiques chromatiques (moyenne, écart type sur la forme). Ces opérations sont classiques dans l'art du traitement d'images.

Comme décrit précédemment, on associe un ensemble de types conceptuels à un ensemble de caractéristiques physiques. Ces types

conceptuels sont par exemple : maison, animal, fleur, etc. Les caractéristiques physiques correspondent aux formes simples et à leurs descriptions chromatiques.

L'étape suivante consiste à rechercher les proximités entre des formes inconnues et des formes « statistiquement » connues (voir plus loin) ou une moyenne formant une « meilleure probabilité » qu'une image nous plaise. Ici, les types conceptuels sont ceux inhérents aux images fixes (on adapte cet ensemble dans le cas des vidéos : action, sport, etc. information de type dynamique). Comme dans le cas de données sonores, la vidéo est considérée comme un échantillonnage de plusieurs images étant traité isolément puis contribuant à une caractérisation statistique du film.

On prend ici en compte une incertitude et une ambiguïté des analyses. Les êtres vivants prennent leurs décisions dans un contexte d'incertitude, en terme de probabilité. Même des décisions qui peuvent nous apparaître comme évidentes, répondent à ce principe.

Lorsque nous apercevons une personne au loin, notre cerveau mobilise ses fonctions cognitives pour comparer ce que nous percevons à un ensemble de caractéristiques de base que nous avons mémorisé dans le passé. Ces opérations nous amènent à tirer la conclusion que la personne en question est probablement un ami. Lorsque nous voyons cette personne de plus près, son identité nous apparaît comme évidente pourtant, les mêmes mécanismes que précédemment nous ont amené à une réponse tout aussi probabiliste. La différence, c'est que dans le second cas, la probabilité est beaucoup plus importante.

On va décrire maintenant deux mécanismes utilisés dans l'exemple préféré de réalisation de l'invention.

Le premier est le traitement et la composition (le calcul) de prémisses approximatives (peut être que, il y a des chances, etc.) La logique floue donne un certain nombre d'outils permettant de calculer avec des éléments incertains. L'autre mécanisme est basé sur le croisement d'information. Par exemple, si l'analyse d'une image révèle une forme proche à 75% d'un camion, à 75% d'une maison, à 60% d'un bus et à 10%

d'un chalet, on considérera que la forme est un véhicule car sémantiquement la catégorie véhicule (camion et bus) est plus « co-désignée » que la catégorie habitation (chalet et maison). C'est un exemple particulier de ce que l'on appelle ici l'inférence corrélatrice. Ceci est possible
5 car les types conceptuels possèdent leur propre réseau d'association correspondant aux caractéristiques linguistiques (arbre, fleur et végétal sont associés mais de manière différente). Au départ, la recherche des caractéristiques et le travail de rapprochement entre ces caractéristiques et les types conceptuels se fait sans tenir compte du niveau d'association
10 interne entre les types conceptuels. Cette association sera utilisée pour lever certaines ambiguïtés sur les associations physiques/conceptuels.

Là encore, les recoupements sémantiques reflètent des affinités entre les différents types, affinités qui sont prises en considération en commun avec les probabilités d'appartenance à chacun de ces types, et
15 permettent d'établir l'appartenance ou non à une catégorie.

En d'autres termes, pour identifier dans cet exemple une catégorie générale d'appartenance, on utilise les probabilités d'appartenance aux différents types ainsi obtenues en combinaison avec une série de relations d'affinités entre types, affinités représentatives de la perception humaine,
20 de manière à élire des appartenances qui sont à la fois majoritaires en probabilité et qui se co-désignent par leur relation d'affinité, et de manière à exclure des appartenances ayant une moindre affinité avec les types élus.

Pour l'établissement des relations d'affinités utilisées dans cette deuxième étape ici appelée « étape d'inférence corrélatrice », on met ici en œuvre un traitement préliminaire d'une multitude de textes, dans lequel on relève les occurrences particulièrement fréquentes de deux termes à
25 chaque fois en position de proximité dans ces textes. Ainsi, la relation sémantique entre « maison » et « chalet » est identifiée simplement par la présence fréquente de ces deux termes proches l'un de l'autre dans une
30 série de textes.

A contrario, le fait que les termes « camion » et « chalet » ne soient pratiquement jamais présents et proches l'un de l'autre, permet d'identifier qu'aucun lien de proximité sémantique (affinité) n'existe entre ces mots.

L'existence d'une affinité est par exemple décelée en fonction du dépassement d'un seuil de fréquence dans les textes de deux mots à proximité l'un de l'autre (à proximité car seulement distants d'un nombre de mots prédéfinis).

5 Un tel traitement préalable d'une série de textes permet d'automatiser l'établissement d'affinités entre types. De telles affinités sont bien représentatives de la perception humaine en ce qu'elles correspondent au sentiment d'affinité entre types que peut ressentir une personne, que ce soit l'affinité entre « rock » et « hard-rock » pour la musique, ou l'affinité
10 entre « maison » et « chalet » pour des images.

Bien entendu, dans d'autres modes de réalisation, le traitement préalable d'une série d'objets descriptifs ou d'une série d'objets existants permet de la même façon d'identifier des affinités typiques car statistiquement très représentées comme définissant une relation
15 sémantique entre deux termes ou deux caractéristiques.

Les élections et exclusions de types consistent simplement ici à identifier une partie d'entre eux signalant ensemble la catégorie générale d'appartenance de l'objet.

La constitution du réseau associatif (figure 1a, plan du bas) est
20 réalisée ici par apprentissage, comme on vient de le décrire à partir de l'exploitation d'une série de textes, dans l'exemple en présence. Maintenant, il s'agit d'identifier les caractéristiques statistiques de la relation entre entrée (physique, plan du haut) et sortie (conceptuel, plan du bas). Ces caractéristiques seront utilisées pour alimenter le modèle flou. Si on
25 considère un ensemble de z relations R entre chacun des p types de référence et une ou plusieurs des n caractéristiques, chaque relation R_{ij} contient la représentation statistique de l'ensemble des exemples de cette relation fournis par les objets de référence. La relation contient la moyenne, l'écart moyen, min, max, n , etc. Par exemple, sur cent objets de référence
30 la caractéristique i du type j est présente soixante fois la moyenne de la relation R_{ij} est à 0.6.

Pour plus de clarté, nous allons donner maintenant un exemple d'application d'une prise de décision à partir d'une approche floue. La reconnaissance de contenu d'image est ici l'exemple d'application.

L'objectif est d'implémenter dans un réseau associatif des relations
5 entre des formes et un type par exemple le type voiture. Nous nous limitons dans le cadre de cet exemple mais dans la pratique on peut prendre plus de caractéristiques que la forme et plus de type que le simple type voiture.

Sur la figure 2, l'étape A correspond à une identification manuelle
10 des caractéristiques de références (e.g. formes, spectre moyen sur les 10 premières secondes d'un morceau de musique) et des types pertinents (travail d'expert).

L'étape B est une récupération automatique (moteur de recherche et
aspirateur Web-Internet) d'objets (par exemple image de voiture)
statistiquement pertinents décrivant un type donné (idem pour tous les
15 types définis) à partir de mots clés (voiture, automobile, etc).

L'étape C est une étape d'analyse du contenu des objets de
références : mesure des caractéristiques de référence, analyse et
segmentation des résultats (groupes). On utilise une technique de
catégorisation classique pour l'identification des groupes.

20 Ces trois étapes sont réalisées pour plusieurs types de références reliés par un réseau associatif. Cette approche permettra dans une étape ultérieure (Etape H), par recoupement, de régler certaines ambiguïtés.

L'étape D correspond à une fuzzification, c'est-à-dire une
identification des variables d'entrée (les caractéristiques) et de sortie (les
25 types), des fonctions et limites d'appartenance, les informations issues des étapes précédentes permettant d'automatiser cette étape.

L'étape E correspond à une étape de définition des règles
d'inférences. Ces règles sont génériques et correspondent à une
composition homogène de fonctions AND/OR liant les entrées aux sorties.

30 L'étape F est une étape de mesure des caractéristiques sur l'objet inconnu.

L'étape G est une étape de défuzzification : évaluation du niveau des
variables de sortie. Cette étape présentée ici au sens global correspond à

la défuzzification de chacune des sorties. Le résultat se présente sous la forme d'une probabilité d'appartenance de l'objet inconnu à chaque type.

L'étape H est une étape de croisement des résultats à l'aide d'un réseau associatif pour limiter l'ambiguïté (e.g. si l'objet est 50% un arbre, 50% une voiture et 50% un camion alors ce n'est certainement pas un arbre).

On va décrire maintenant, de manière plus détaillée, ce synoptique de la figure 2, dans une première étape de définition de références (étapes A à C sur la figure 2), on récupère un nombre important d'images contenant des voitures et on identifie les contours des formes à l'aide d'outils logiciels classiques. Ces formes sont normalisées (ramenées à des proportions identiques) et comparées (repère cartésien) de manière à identifier des groupes homogènes (contour superposable avec erreur minimale). Pour chacun de ces groupes on identifie la forme moyenne, dite de référence (qui peut être une image quelconque du groupe, puisque la segmentation – étape C – assure que les membres d'un groupe ont une forte similitude) et le nombre de candidats par groupe. On pose que les formes de références regroupant le plus de candidats décrivent le mieux de type de référence (voiture). On identifie le nombre optimal de groupes avec des techniques classiques de clustering. Par exemple, on optimise le rapport entre erreur moyenne intra groupes et inter groupes. L'algorithme de clustering est alimenté par la matrice des erreurs forme à forme.

On suppose pour la suite de l'exemple que le nombre de groupes idéal est deux. On dispose donc de deux formes moyennes dont chacune affectée à une probabilité de similitude fonction du % de forme dans le groupe par rapport au nombre total de formes.

Contrairement à la logique binaire (une proposition est vraie ou fausse), la logique floue prend en compte le caractère « possible » d'une proposition en y associant un certain niveau de vraisemblance. En plus d'un concept, la logique floue introduit des méthodes et une algèbre adaptée permettant de manipuler des grandeurs floues (approximatives).

La première étape (étape D sur la figure 2) dite de quantification floue (ou « fuzzification ») consiste à modéliser chaque entrée du système

par des courbes donnant le degré d'appartenance aux différents états du système possiblement associé à ces entrées. Dans notre cas, les entrées sont les caractéristiques physiques et les états sont les types (ou associations de types) conceptuels. Le système étant le réseau d'association (pondération normalisée) existant entre les entrées et les sorties (voir figure 1a).

Ainsi, chaque entrée, en fonction de son niveau de pondération sera représentée sur une courbe de type sigmoïde (figure 3 ou approximé figure 4) qui explicitera son niveau d'appartenance à un type donné.

10 Soit par exemple un réseau (figure 1) où le niveau de pondération maximal (normalisé) est de 100. Une caractéristique physique qui sera reliée à un type conceptuel avec une pondération de 10 sera considérée comme la décrivant peu probablement. Au contraire, une pondération de 80 sera considérée comme une description très pertinente.

15 Dans la figure 4, les valeurs min et max obtenues par analyse statistique, fournissent les seuils au-delà desquels la caractéristique est considérée ne plus décrire du tout de type (min) ou très bien le décrire (max).

20 La fonction d'appartenance représentée figure 5 établit le niveau d'appartenance à chaque sous-ensemble flou en fonction du niveau de l'erreur entre les formes de référence (les deux formes moyennes) et les formes contenues dans l'image à analyser.

25 Sur la figure 5, le symbole « - » représente une appartenance peu probable. Le symbole « = » représente une appartenance possible. Le symbole « + » représente une appartenance très probable.

30 Sim_{min} et Sim_{max} désignent respectivement la similitude minimale et maximale calculées forme à forme dans chaque groupe lors de l'apprentissage. Sim_{max} est la similitude entre les deux formes les plus semblables et Sim_{min} la similitude entre les deux formes les plus différentes au sein d'un groupe donné.

D'une manière similaire, on définit la fonction d'appartenance pour la variable de sortie (figure 6).

Les valeurs Sim_{min} et Sim_{max} décrivent ici l'incertitude minimale et maximale pour la prise de décision. On dira qu'au dessous de Sim_{min} la forme n'appartient pas au type, et qu'au dessus de Sim_{max} , elle appartient au type et qu'entre ces deux valeurs elle appartient au type avec un certain

5 niveau de probabilité. Par exemple, je peux considérer qu'en dessous de 20% de certitude la personne que je vois au loin n'est pas la personne que j'imagine, alors qu'au-dessus de 75%, j'estime que c'est la personne en question.

Les règles d'inférence (étapes E et F sur la figure 2 : définition des

10 règles d'inférence et mesure), impliquent une suite d'opérations logiques et des conditions liant les entrées et la sortie que l'on peut adapter plus ou moins finement selon les cas à traiter. Le tableau suivant indique les modes de composition liant les deux groupes pour prendre la décision d'appartenance ou pas au type de référence. Il faut rappeler ici que les

15 deux groupes, obtenus par segmentation (étape C) sont des groupes dissimilaires. Une appartenance aux deux groupes ne permet donc pas de lever l'ambiguïté concernant l'appartenance au type. C'est pour cette raison que l'on affecte un – dans le tableau lorsque E1 et E2 sont à +.

Les règles d'inférence entre les entrées (le niveau d'appartenance

20 au groupe 1 ou 2) et la sortie (niveau de certitude qu'il s'agit d'une voiture) sont les suivantes. Les deux entrées étant reliées à la sortie par une fonction OR (c'est à dire le maximum des deux valeurs en logique floue) dans le cas des faibles probabilités et de la fonction AND pour les fortes.

E1/E2	-	=	+
-	-	-	=
=	-	=	+
+	=	+	-

25

Table 1 : inférences liant les entrées à la sortie

Ce tableau s'interprète comme suit :

Si (E1+ et E2+) ou (E1- et E2-) ou (E1- et E2=) ou (E1= et E2-) alors

S-

(Si la ressemblance à la forme 1 et à la forme 2 est faible ou que la ressemblance est faible à la forme 1 et médiocre à la forme 2 ou que la ressemblance est médiocre à la forme 1 et faible à la forme 2 alors la forme inconnue est peu probablement une voiture.

- 5 Si (E1= et E2=) ou (E1+ et E2-) ou (E1- et E2+) alors S=
 Si (E1+ et E2=) ou (E1= et E2+) alors S+

On donne ici maintenant un exemple d'algorithme de composition des règles d'inférences pour un nombre de groupes quelconques.

- 10 Chaque caractéristique peut prendre trois états (+, -, =) soit, pour n caractéristiques, 3^n combinaisons possibles.

Dans cette étape, on initialise les valeurs S+ ; S- ; et S = à 0.

Pour chacune des 3^n combinaisons :

- on identifie le signe et la valeur des éléments de la combinaison ;
 - on calcule Min comme étant le minimum de tous les éléments de la
- 15 combinaison.

- dans chaque combinaison, on compte le nombre + et le nombre -.

Les résultats sont alors exploités de la manière suivante :

- si le nombre d'éléments « + » est inférieur à 30% du nombre d'éléments de la combinaison et au moins égal à 1 élément (on prend la
- 20 valeur entière la plus proche des 30% (ex : si 0,7 alors on prend 1), alors, si Min est supérieur à la valeur S+, la valeur S+ est définie comme étant égale à Min ;

- si le nombre d'éléments est supérieur à 70 %, alors, si Min est supérieur à la valeur S =, la valeur de S = est définie comme étant égale à
- 25 min ;

- si le nombre d'éléments est compris entre 30 et 70%, alors, si Min est supérieur à la valeur S =, la valeur S = est définie comme étant égale à Min)

- 30 La sortie de ce système (étapes G et H : defuzzification et croisement) est de la forme : l'objet correspond (par exemple) à 80% au type T1, à 65% au type T2, etc. Dans notre exemple, nous nous sommes limités à un seul type T1=voiture.

La defuzzification consiste à évaluer la grandeur de sortie (ici la probabilité d'appartenance à un type donné). Ceci est réalisé en projetant l'état des variables d'entrées sur les variables de sortie au travers des règles d'inférence. Pratiquement, on réduit (décapite) les fonctions d'appartenance des variables de sortie suivant les résultats fournis par les opérations logiques (floues) dictées par les règles d'inférences. Le paragraphe suivant donne un exemple illustré de cette opération.

On donne maintenant un exemple d'application. On considère qu'après la phase d'apprentissage, chaque groupe a les caractéristiques suivantes (figures 7 et 8) :

Similitude /groupe	1	2
Similitude -min %	18	25
Similitude -max %	58	62

On considère qu'une image possède une forme qui après comparaison avec les deux formes de références a les erreurs suivantes :

Similitude /groupe	1	2
Sim	20	45

15

Si (E1+ et E2+) ou (E1- et E2-) ou (E1- et E2=) ou (E1= et E2-) alors S-
 $S- = \text{Max} [\text{Min} (0, 0.25), \text{Min}(0.85, 0), \text{Min}(0.85, 0.7), \text{Min}(0.3, 0)] = 0.7$

Si (E1= et E2=) ou (E1+ et E2-) ou (E1- et E2+) alors S=
 $S= = \text{Max} [\text{Min}(0.3, 0.25), \text{Min}(0, 0), \text{Min}(0.85, 0.25)] = 0.25$

20 Si (E1+ et E2+) (E1+ et E2=) ou (E1= et E2+) alors S+
 $S+ = \text{Max} [\text{Min}(0, 0.7), \text{Min}(0.3, 0.25)] = 0.25$

Le résultat 48 correspond au centre de gravité des trois ensembles (-, +, =) de la sortie réduit respectivement à 0.7, 0.25, 0.25 (surface hachurée, figure 9)

25 La réponse est donc : « c'est une voiture à 48% ». Puisque dans notre exemple, nous nous sommes limités à un seul type, nous ne pouvons pas faire de croisements.

Dans la réalité, on met en œuvre en plus l'étape consistant à utiliser les probabilités d'appartenance aux différents types ainsi obtenues en combinaison avec une série de relations d'affinités entre types, affinités représentatives de la perception humaine, de manière à élire des appartenances qui sont à la fois majoritaires en probabilité et qui se co-désignent par leur relation d'affinité, et de manière à exclure des appartenances ayant une moindre affinité avec les types élus.

Cette étape, en combinaison avec les étapes précédentes, permet de multiplier notablement la précision de l'analyse. Globalement, cette méthode est intéressante car elle offre une loi de composition pour des entrées floues ce qui est d'autant plus difficile à gérer qu'il y a de nombreuses entrées.

REVENDEICATIONS

1. Procédé de description automatique d'un objet multimédia
5 inconnu, dans lequel on associe l'objet inconnu avec plusieurs types
d'objets multimédia de référence selon à chaque fois une probabilité
d'appartenance respective à chaque type considéré (G), le procédé incluant
une étape consistant à mesurer au moins une caractéristique physique sur
l'objet inconnu (F) et à la comparer avec des mesures de caractéristiques
10 représentant au mieux les types de référence, le procédé comprenant en
outre l'étape (H) consistant à utiliser pour chaque type au moins une
relation probabiliste donnant une probabilité d'appartenance au type
considéré en fonction du résultat de la comparaison des caractéristiques de
l'objet inconnu et du type, le procédé comprenant en outre l'étape
15 consistant à utiliser les probabilités d'appartenance aux différents types
ainsi obtenues en combinaison avec une série de relations d'affinités entre
types, de manière à élire des appartenances qui sont à la fois majoritaires
en probabilité et qui se co-désignent par leur relation d'affinité, et de
manière à exclure des appartenances ayant une moindre affinité avec les
20 types élus.

2. Procédé de description selon la revendication 1, caractérisé en ce
qu'il comporte l'étape préliminaire consistant à définir les caractéristiques
de référence d'un type à partir d'un groupe d'objets multimédia sensés
représenter ce type (B), en réalisant une mesure d'une caractéristique
25 physique sur l'ensemble de ce groupe, et en tirant de cette série de
mesures une ou des valeur(s) de référence pour cette caractéristique,
valeur(s) de référence qui est(sont) ensuite utilisée(s) pour définir la relation
de probabilité d'appartenance au type en tant que valeur à laquelle on
compare une mesure sur un objet inconnu pour déduire la probabilité
30 d'appartenance au type.

3. Procédé selon la revendication 2, caractérisé en ce que le groupe
d'objets multimédia est fourni à l'aide d'une étape de recherche

automatique dans un système d'information avec un aspirateur internet et un moteur de recherche internet.

4. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il comporte l'étape préliminaire consistant à mesurer
5 sur un ensemble d'objets descriptifs une fréquence d'occurrence simultanée des types dans ces objets et en déduire l'existence d'une affinité entre au moins deux types lorsque ceux-ci ont une occurrence simultanée particulière.

5. Procédé de description selon l'une quelconque des revendications
10 précédentes, caractérisé en ce qu'il comporte l'étape consistant à effectuer une reconnaissance de forme sur l'objet multimédia inconnu (F), au moins une forme de référence à reconnaître sur l'objet inconnu constituant une caractéristique physique appartenant à la définition d'un des types.

6. Procédé selon la revendication 5, caractérisé en ce que la
15 reconnaissance de forme comprend une mesure de similitude entre une forme prélevée sur l'objet et la forme de référence, et comprend également l'utilisation d'une relation prédéfinie qui donne une probabilité d'appartenance (G) au type en fonction de la mesure de similitude de forme effectuée.

20 7. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'au moins un type inclut plusieurs caractéristiques de référence (E1, E2), en ce qu'on réalise au moins deux mesures sur l'objet inconnu pour établir une mesure de proximité avec chacune des deux caractéristiques de référence, et en ce que l'on établit la probabilité
25 d'appartenance à ce type en utilisant au moins deux relations, chacune donnant une probabilité d'appartenance au type en fonction de la proximité à une caractéristique différente, et en ce que les deux relations de probabilité sont utilisées pour établir un résultat global de probabilité d'appartenance de l'objet au type considéré.

30 8. Procédé selon la revendication précédente, caractérisé en ce que les au moins deux relations de probabilité d'appartenance au type sont exploitées selon une technique combinatoire de logique floue pour fournir le résultat de probabilité d'appartenance de l'objet au type considéré.

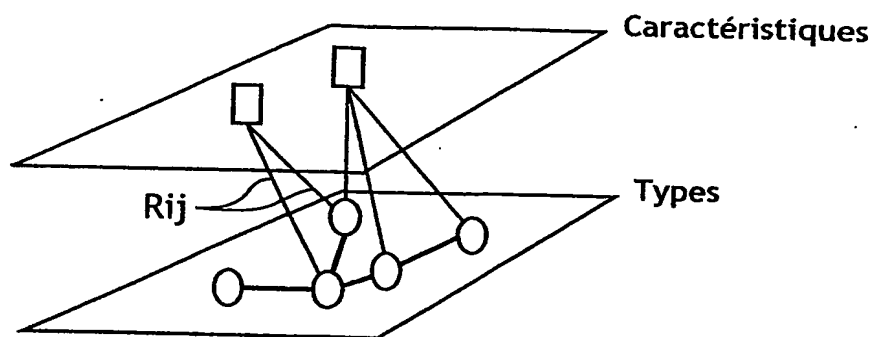
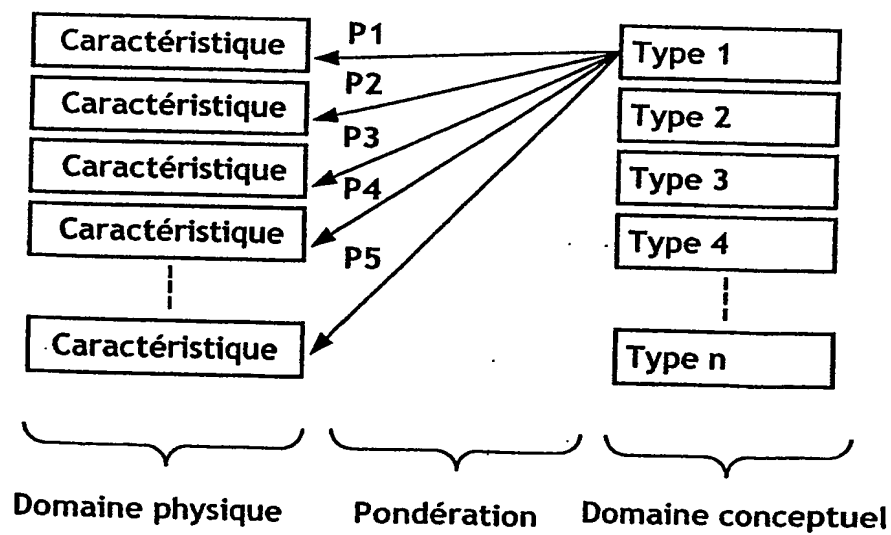
9. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce qu'on met en œuvre une technique de logique floue qui consiste en un mécanisme donnant un seul niveau de probabilité d'appartenance à un type de référence à partir d'une combinaison de
5 probabilités de similitude aux différentes caractéristiques du type de référence.

10. Dispositif de description automatique d'un objet multimédia inconnu, comprenant des moyens pour associer l'objet inconnu avec plusieurs types d'objets multimédia de référence selon à chaque fois une
10 probabilité d'appartenance respective au type considéré (G), le dispositif incluant des moyens pour mesurer au moins une caractéristique physique sur l'objet inconnu (F) et la comparer avec des mesures représentant au mieux les types de référence, le dispositif comprenant en outre des moyens utilisant pour chaque type au moins une relation probabiliste donnant une
15 probabilité d'appartenance au type considéré en fonction du résultat de la comparaison des caractéristiques de l'objet inconnu et du type, le dispositif comprenant en outre des moyens pour utiliser les probabilités d'appartenance aux différents types ainsi obtenues en combinaison avec une série de relations d'affinités entre types, de manière à élire (H) des
20 appartenances qui sont à la fois majoritaires en probabilité et qui se co-désignent par leur relation d'affinité, et de manière à exclure des appartenances ayant une moindre affinité avec les types élus.

11. Dispositif selon la revendication 10, caractérisé en ce qu'il inclut des moyens de traitement pour exploiter plusieurs groupes d'objets
25 multimédia de référence (B, C), chaque groupe représentant au mieux son type correspondant, ces moyens de traitement étant en outre prévus pour réaliser au moins une mesure d'une caractéristique physique sur l'ensemble d'un groupe considéré, et en tirer une mesure de référence de cette caractéristique, cette mesure de référence étant ensuite utilisée dans
30 la définition de la relation donnant une probabilité d'appartenance au type considéré, en tant que mesure à laquelle le dispositif compare une mesure sur un objet inconnu pour déduire la probabilité d'appartenance au type considéré (G).

12. Dispositif selon la revendication 10 ou la revendication 11, caractérisé en ce qu'il inclut des moyens pour mettre en œuvre une étape préliminaire consistant à mesurer sur un ensemble d'objets descriptifs une fréquence d'occurrence des types dans ces objets, et en déduire l'existence
- 5 d'une affinité entre au moins deux types lorsque ceux-ci ont une occurrence simultanée particulière.

1/5

FIG.1aFIG.1b

2/5

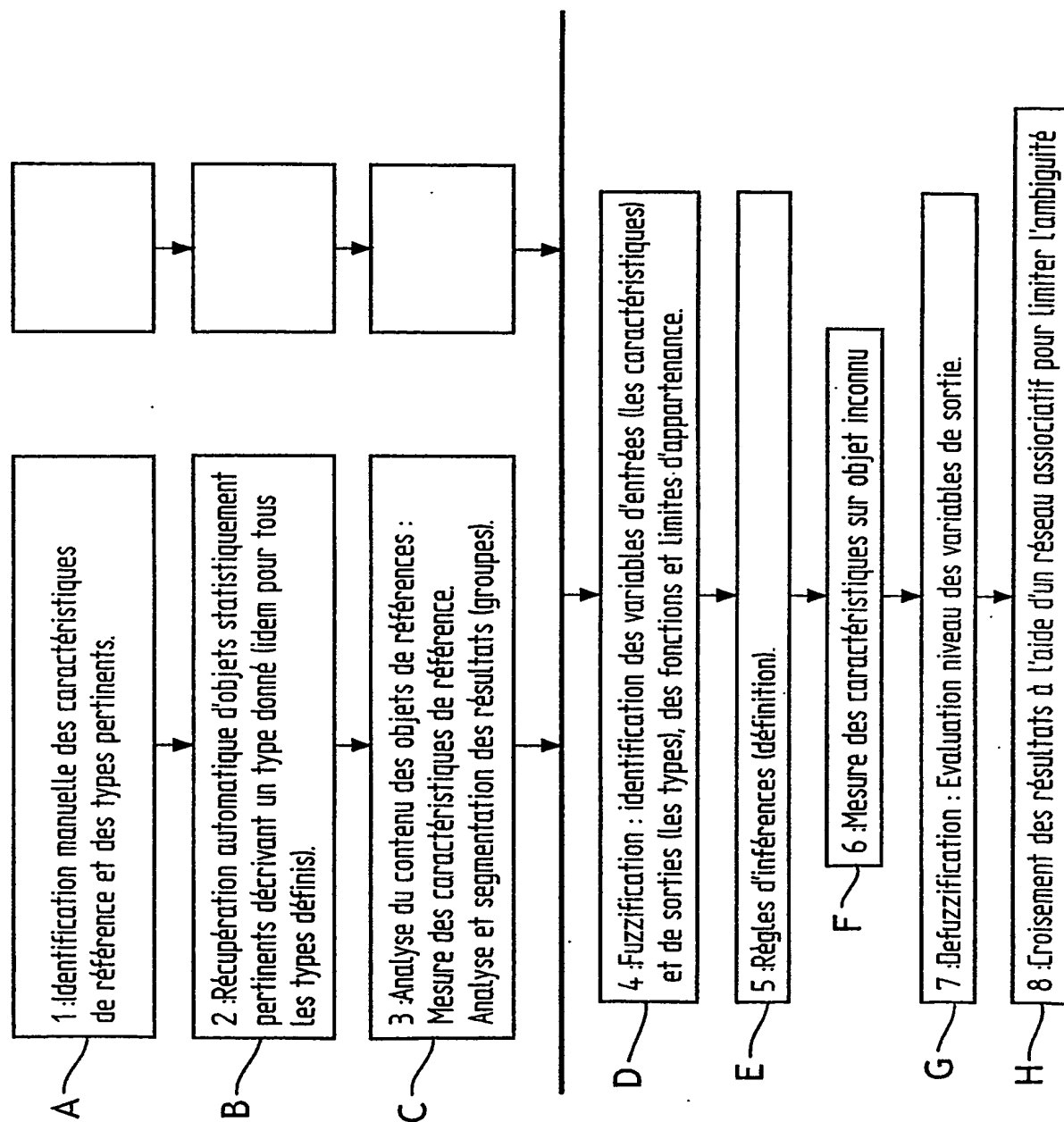
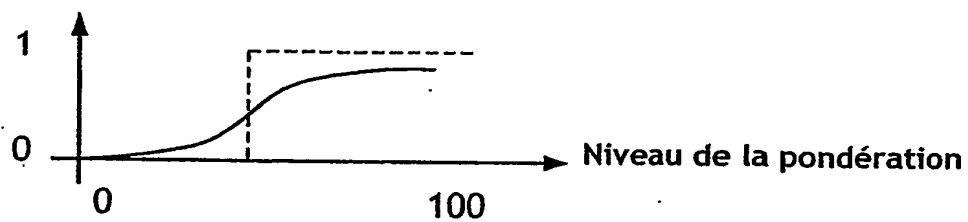
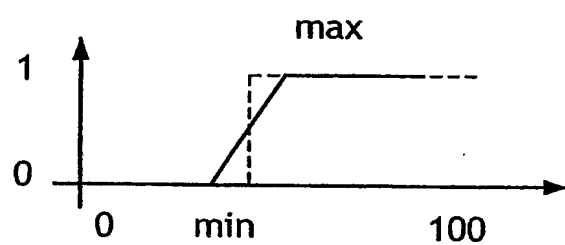
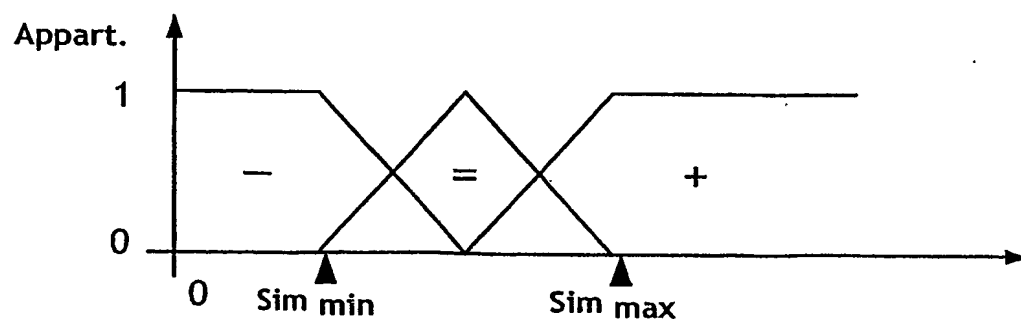


FIG.2

3/5

FIG.3FIG.4FIG.5

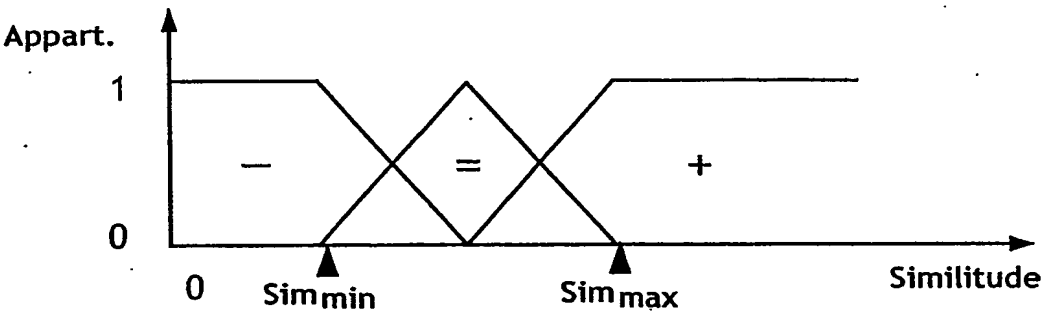


FIG.6

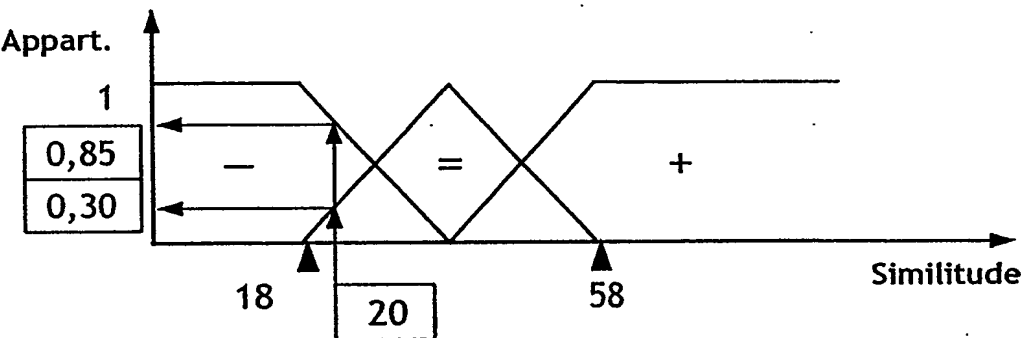


FIG.7

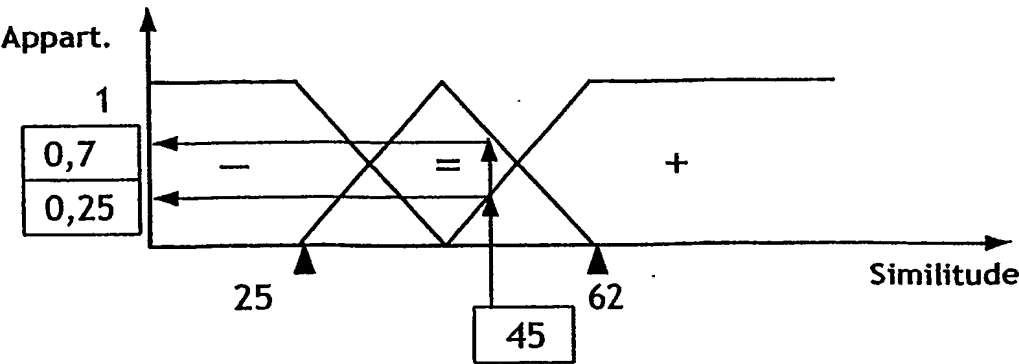
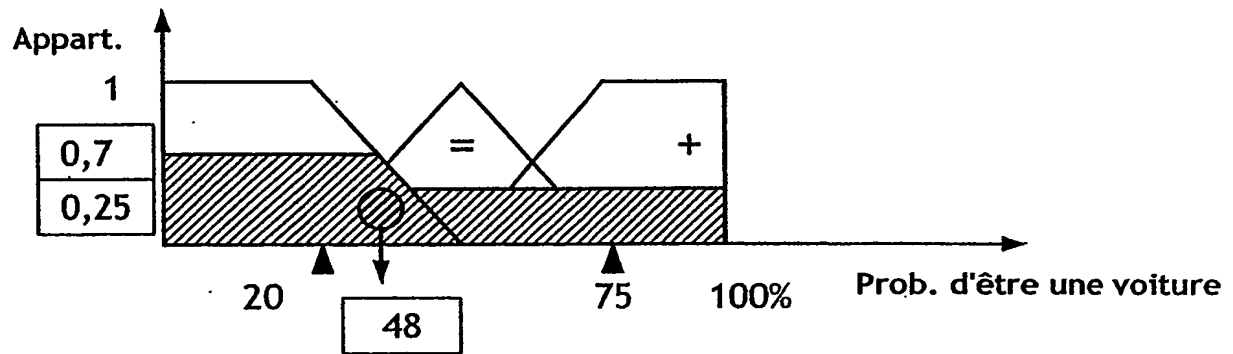


FIG.8

5/5

FIG.9